

Quantized Consensus by the ADMM: Probabilistic versus Deterministic Quantizers

Shengyu Zhu, *Student Member, IEEE*, and Biao Chen, *Fellow, IEEE*

Abstract—This paper develops efficient algorithms for distributed average consensus with quantized communication using the alternating direction method of multipliers (ADMM). We first study the effects of probabilistic and deterministic quantizations on a distributed ADMM algorithm. With probabilistic quantization, this algorithm yields linear convergence to the desired average in the mean sense with a bounded variance. When deterministic quantization is employed, the distributed ADMM either converges to a consensus or cycles with a finite period after a finite-time iteration. In the cyclic case, local quantized variables have the same mean over one period and hence each node can also reach a consensus. We then obtain an upper bound on the consensus error which depends only on the quantization resolution and the average degree of the network. Finally, we propose a two-stage algorithm which combines both probabilistic and deterministic quantizations. Simulations show that the two-stage algorithm, without picking small algorithm parameter, has consensus errors that are typically less than one quantization resolution for all connected networks where agents' data can be of arbitrary magnitudes.

Index Terms—Quantized consensus, dither, probabilistic quantization, deterministic quantization, alternating direction method of multipliers, linear convergence.

I. INTRODUCTION

IN recent years there has been considerable interest in distributed average consensus where a group of agents aim to reach a consensus on the average of their measurements [2]–[20]. This is largely motivated by numerous applications in control, signal processing, and computer science. For example, the distributed averaging is a fundamental problem in *ad hoc* network applications, such as distributed agreement and synchronization [4], distributed coordination of mobile autonomous agents [5], and distributed data fusion in sensor networks [6]. It has also found applications in load balancing for parallel computers [7].

We consider in this paper distributed averaging algorithms where nodes only exchange information with their immediate neighbors. These algorithms are extremely attractive for large scale networks characterized by the lack of centralized access to information. They are also energy efficient and enhance the survivability of the networks, compared with fusion center based processing. However, a number of factors such as limited bandwidth, sensor battery power, and computing resources

place tight constraints on the rate and form of information exchange amongst neighboring nodes, resulting in *quantized consensus* problems [2], [8]. This paper is specifically devoted to developing efficient algorithms for quantized consensus in connected networks with static topologies.

A. Related work

There are three widely used methods for solving distributed averaging problems. A classical approach is to update the state of each node with a weighted average of values from neighboring nodes [9]–[11]. The matrix, consisting of the weights associated with the edges, is chosen to be doubly stochastic to ensure convergence to the average. Another method is a gossip based algorithm, initially introduced in [21] for consensus problems and further studied in [8], [12], [13], among others. The third approach is to employ the ADMM which is an iterative algorithm for solving convex problems and has received much attention recently (see [22] and references therein). The idea is to formulate the data average as the solution to a least-squares problem and manipulate the ADMM updates to derive a distributed algorithm [14]–[16].

In the most ideal case where agents are able to send and receive real values with infinite precision, the three methods all lead to the desired consensus at the average. When quantization is imposed, however, these methods do not directly apply. A well studied approach for quantized consensus is to use dithered quantizers which add noises to agents' variables before quantization [23]. By imposing certain conditions, the quantization error sequence becomes independent and identically distributed (i.i.d.) and is also independent of the input sequence. The classical approach and the gossip based algorithm then yield the almost sure consensus at a common but random quantization level with the expectation of the consensus value equal to the desired average [17], [18], [20]. To the best of our knowledge, there have been no existing results on the ADMM based method for quantized consensus. Nevertheless, since the quantization error of dithered quantizer is zero-mean and has a bounded variance, we can immediately extend the results in [15], [16] to quantized consensus (see Section IV). That is, the ADMM based method using dithered quantization leads to the consensus at the data average in the mean sense whose variance converges to a finite value.

Meanwhile, studies on distributed average consensus with deterministic quantizers have been scarcely reported. Deterministic quantization makes the problem much harder to deal with as the error terms caused by quantization no longer possess tractable statistical characteristics [17], [18]. The authors

This work was supported by National Science Foundation under Award CCF1218289, by Army Research Office under Award W911NF-12-1-0383, and Air Force Office of Scientific Research under Award FA9550-10-1-0458. The material of this paper was presented in part at IEEE GlobalSIP 2015 [1].

S. Zhu and B. Chen are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: szhu05@syr.edu; bichen@syr.edu).

in [11] show that the classical approach, where a quantization rule that rounds the values down is adopted, converges to a consensus with an error from the average depending on the quantization resolution, the number of agents, the agents' data and the updated weights of each agent. A recent result of [19] indicates that this approach, with appropriate choices of the weights, reaches a quantized consensus close to the average in finite time or leads all agents' variables to cycle in a small neighborhood around the average; in the latter case, however, the consensus is not guaranteed. The gossip based algorithms in [20] and [8] have similar results to those of the classical approach. The ADMM based algorithms for deterministically quantized consensus, however, have not yet been explored.

B. Our contributions

One shall note that the consensus error for deterministically quantized consensus in [11], [20] is much undesired when the number of agents or the range of agents' data becomes very large. Unfortunately, this is typically the case in large scale networks or big data settings. The ADMM has been known to be an efficient algorithm for large scale optimizations and used in various applications such as regression and classification [22]. Moreover, [24]–[26] validate the fast convergence of the ADMM and [15], [16] demonstrate the resilience of the ADMM to noise, link failures, etc. We therefore expect ADMM based methods to work well for quantized consensus problems, with regards to both the consensus error and the convergence time.

We first study the effect of probabilistic quantization [27], which is equivalent to a dithering method as shown by [17, Lemma 2], on the ADMM based method. Utilizing the first and second order moments of the probabilistic quantizer output, we establish the convergence to the average in the mean sense based on existing convergence results of the ADMM. Furthermore, recent work of [28] enables us to immediately characterize the convergence rate of the distributed ADMM with probabilistic quantization.

The main contribution of this paper is to design and analyze an ADMM based approach using deterministic quantization. We establish that a distributed deterministically quantized ADMM algorithm either converges to a consensus or cycles around the average after a finite-time iteration as long as a mild initialization condition is satisfied. We also show that the cyclic period is finite and that the quantized variable at each node has the same mean over one period. Thus, a consensus can be reached within finite iterations for both convergent and cyclic cases. We then derive an upper bound that only depends on the quantization resolution and the average degree of the undirected graph (two times the ratio of the number of edges to the number of nodes). This is much preferred for large scale networks as it does not rely on the number of agents or the agent's data.

While numerical examples show that the deterministically quantized ADMM converges in most cases, we notice that it may reach different consensus values with different initial variable values. It is well known that a good starting point usually helps in such settings. This inspires our approach

for quantized consensus which first uses the probabilistic method to obtain a good starting point and then employs the deterministic algorithm. Simulations show that this two-stage approach tends to converge and also performs best among all existing methods using deterministic quantization in terms of the consensus error.

C. Paper organization

The rest of this paper is organized as follows. Section II reviews the application of the ADMM to the distributed averaging problem without quantization, which leads to a distributed ADMM algorithm. We then develop several convergence results of this algorithm; they will be used later to establish our main results. Section III defines probabilistic and deterministic quantization schemes. Their effects on the distributed ADMM are studied respectively in Sections IV and V. Section VI describes the proposed algorithm for quantized consensus which combines the two quantized ADMM methods, followed by simulation results in Section VII. Section VIII concludes the paper.

D. Notations

Denote by $\|x\|_2$ the Euclidean norm of a vector x and $\langle x, y \rangle$ the inner product of two vectors x and y . Given a semidefinite matrix G with proper dimensions, the G -norm of x is $\|x\|_G = \sqrt{x^T G x}$. Also denote $\sigma_{\max}(M)$ as the largest singular value of a square matrix M and $\tilde{\sigma}_{\min}(M)$ as the smallest nonzero singular value of M .

We use two definitions of rate of convergence for an iterative algorithm. A sequence x^k , where the superscript k stands for time index, is said to converge Q -linearly to a point x^* if there exists a number $v \in (0, 1)$ such that $\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = v$ with $\|\cdot\|$ being a vector norm. A sequence y^k is said to converge R -linearly to y^* if for all k , $\|y^k - y^*\| \leq \|x^k - x^*\|$ where x^k converges Q -linearly to x^* .

II. DISTRIBUTED AVERAGE CONSENSUS BY THE ADMM

This section introduces the consensus ADMM (CADMM) for average consensus without quantization. This ideal case provides a good understanding of how the ADMM works for distributed average consensus. We start with the setting of the distributed averaging problem.

A. Problem setting

Consider a connected network of N agents which are bidirectionally connected by E edges (and thus $2E$ arcs). We describe this network as a symmetric directed graph $\mathcal{G}_d = \{\mathcal{V}, \mathcal{A}\}$ or an undirected graph $\mathcal{G}_u = \{\mathcal{V}, \mathcal{E}\}$, where \mathcal{V} is the set of vertices with cardinality $|\mathcal{V}| = N$, \mathcal{A} is the set of arcs with $|\mathcal{A}| = 2E$ and \mathcal{E} is the set of edges with $|\mathcal{E}| = E$. Assume that the topology of the network is fixed throughout this paper. Let $r_i \in \mathbb{R}$ be the local data only available at node i , $i = 1, 2, \dots, N$, and $\mathbf{r} \in \mathbb{R}^N$ the vector concatenating all r_i . The goal of distributed average consensus is to compute the data average

$$x_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N r_i \quad (1)$$

by local data exchanges among neighboring nodes.

B. Application of the ADMM to distributed average consensus: CADMM

The ADMM applies in general to the convex optimization problem in the form of

$$\begin{aligned} & \underset{\mathbf{y}_1, \mathbf{y}_2}{\text{minimize}} && g_1(\mathbf{y}_1) + g_2(\mathbf{y}_2) \\ & \text{subject to} && \mathbf{C}_1 \mathbf{y}_1 + \mathbf{C}_2 \mathbf{y}_2 = \mathbf{c}, \end{aligned} \quad (2)$$

where \mathbf{y}_1 and \mathbf{y}_2 are optimization variables, g_1 and g_2 are convex functions, and $\mathbf{C}_1 \mathbf{y}_1 + \mathbf{C}_2 \mathbf{y}_2 = \mathbf{c}$ is a linear constraint on \mathbf{y}_1 and \mathbf{y}_2 . The ADMM solves a sequence of subproblems involving g_1 and g_2 one at a time and iterate to converge when, e.g., g_1 and g_2 are proper closed convex functions and the Lagrangian of (2) has a saddle point [22].

To apply the ADMM, we first formulate (1) as a convex optimization problem

$$x_{\text{avg}} = \arg \min_{\tilde{x}} \sum_{i=1}^N \frac{1}{2} (\tilde{x} - r_i)^2, \quad (3)$$

that is, the data average is the solution to a least-squares minimization problem. We continue to reformulate (3) in the form of (2) as

$$\begin{aligned} & \underset{\{x_i\}, \{z_{ij}\}}{\text{minimize}} && \sum_{i=1}^N \frac{1}{2} (x_i - r_i)^2 \\ & \text{subject to} && x_i = z_{ij}, x_j = z_{ij}, \forall (i, j) \in \mathcal{A}, \end{aligned} \quad (4)$$

where x_i is the local copy of the common optimization variable \tilde{x} at node i and z_{ij} is an auxiliary variable imposing the consensus constraint on neighboring nodes i and j . We emphasize that throughout the entire paper, r_i represents the local data, i.e., the observation at the i th agent, while x_i is referred to as the local variable. Since the network is connected, this constraint ensures the consensus to be achieved over the entire network, i.e., $x_i = x_j, \forall (i, j) \in \mathcal{A}$, which in turn guarantees the solution to (4) is the data average x_{avg} . Further define $\mathbf{x} \in \mathbb{R}^N$ as a vector concatenating all x_i , $\mathbf{z} \in \mathbb{R}^{2E}$ as a vector concatenating all z_{ij} , and

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \mathbf{r}\|_2^2. \quad (5)$$

Then (4) can be written in a matrix form as

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{z}}{\text{minimize}} && f(\mathbf{x}) + g(\mathbf{z}) \\ & \text{subject to} && \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{0}, \end{aligned} \quad (6)$$

where $g(\mathbf{z}) = 0$, and $\mathbf{0}$ is a column vector with proper dimensions and all entries being 0. Here $\mathbf{B} = [-\mathbf{I}_{2E}; -\mathbf{I}_{2E}]$ with \mathbf{I}_{2E} being a $2E \times 2E$ identity matrix and $\mathbf{A} = [\mathbf{A}_1; \mathbf{A}_2]$ with $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{2E \times N}$. If $(i, j) \in \mathcal{A}$ and z_{ij} is the q th entry of \mathbf{z} , then the (q, i) th entry of \mathbf{A}_1 and the (q, j) th entry of \mathbf{A}_2 are 1; otherwise the corresponding entries are 0.

We are now ready to apply the ADMM to solve the consensus problem. The augmented Lagrangian of (6) is

$$L_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}\|_2^2, \quad (7)$$

where $\boldsymbol{\lambda} = [\boldsymbol{\beta}; \boldsymbol{\gamma}]$ with $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^{2E}$ is the Lagrange multiplier and ρ is a positive algorithm parameter. At iteration $k+1$, the ADMM first obtains \mathbf{x}^{k+1} by minimizing $L_\rho(\mathbf{x}, \mathbf{z}^k, \boldsymbol{\lambda}^k)$, then calculates \mathbf{z}^{k+1} by minimizing $L_\rho(\mathbf{x}^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k)$ and finally updates $\boldsymbol{\lambda}^{k+1}$ using \mathbf{x}^{k+1} and \mathbf{z}^{k+1} . The updates are

$$\begin{aligned} \mathbf{x}\text{-update} : & \quad \nabla f(\mathbf{x}^{k+1}) + \mathbf{A}^T \boldsymbol{\lambda}^k \\ & \quad + \rho \mathbf{A}^T (\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k) = \mathbf{0}, \\ \mathbf{z}\text{-update} : & \quad \mathbf{B}^T \boldsymbol{\lambda}^k + \rho \mathbf{B}^T (\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1}) = \mathbf{0}, \\ \boldsymbol{\lambda}\text{-update} : & \quad \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k - \rho (\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1}) = \mathbf{0}, \end{aligned} \quad (8)$$

where $\nabla f(\mathbf{x}^{k+1}) = \mathbf{x}^{k+1} - \mathbf{r}$ is the gradient of f at \mathbf{x}^{k+1} .

A nice property of the ADMM, known as *global convergence*, states that the sequence $(\mathbf{x}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)$ generated by (8) has a single limit point $(\mathbf{x}^*, \mathbf{z}^*, \boldsymbol{\lambda}^*)$ which is a primal-dual solution to (7). Proofs can be found in [22], [24], [26]. Noting that our objective function $f(\mathbf{x})$ given in (5) is strongly convex in \mathbf{x} , we obtain $\mathbf{x}^* = \mathbf{1}x_{\text{avg}}$ as the unique primal solution where $\mathbf{1}$ denotes the N -dimensional column vector with all entries being 1. To summarize, we have

Lemma 1 (Global convergence of the ADMM [22], [24], [26]): For any initial values $\mathbf{x}^0 \in \mathbb{R}^N$, $\mathbf{z}^0 \in \mathbb{R}^{2E}$ and $\boldsymbol{\lambda}^0 \in \mathbb{R}^{4E}$, the updates in (8) yield that as $k \rightarrow \infty$,

$$\mathbf{x}^k \rightarrow \mathbf{x}^*, \mathbf{z}^k \rightarrow \mathbf{z}^*, \text{ and } \boldsymbol{\lambda}^k \rightarrow \boldsymbol{\lambda}^*,$$

where $(\mathbf{x}^*, \mathbf{z}^*, \boldsymbol{\lambda}^*)$ is a primal-dual solution to (7) and $\mathbf{x}^* = \mathbf{1}x_{\text{avg}}$ is unique for the distributed average consensus problem (3).

While (8) provides an efficient centralized algorithm to solve (3), it is not clear whether (8) can be carried out in a distributed manner, i.e., data exchanges only occur within neighboring nodes. Interestingly, Lemma 1 states that convergence for the ADMM is guaranteed regardless of initial values $\mathbf{x}^0, \mathbf{z}^0$ and $\boldsymbol{\lambda}^0$; there indeed exist initial values that decentralize (8). Define $\mathbf{M}_+ = \mathbf{A}_1^T + \mathbf{A}_2^T$ and $\mathbf{M}_- = \mathbf{A}_1^T - \mathbf{A}_2^T$ which are respectively the unoriented and oriented incidence matrices with respect to the directed graph \mathcal{G}_d . Initialize $\boldsymbol{\beta}^0 = -\boldsymbol{\gamma}^0$ and $\mathbf{z}^0 = \frac{1}{2} \mathbf{M}_+^T \mathbf{x}^0$. As shown in [28], the updates in (8) lead to

$$\begin{aligned} x_i^{k+1} &= \frac{1}{1 + 2\rho|\mathcal{N}_i|} \left(\rho|\mathcal{N}_i| x_i^k + \rho \sum_{j \in \mathcal{N}_i} x_j^k - \alpha_i^k + r_i \right), \\ \alpha_i^{k+1} &= \alpha_i^k + \rho \left(|\mathcal{N}_i| x_i^{k+1} - \sum_{j \in \mathcal{N}_i} x_j^{k+1} \right) \end{aligned} \quad (9)$$

at node i , where \mathcal{N}_i denotes the set of neighbors of node i and α_i^k is the i th entry of $\boldsymbol{\alpha}^k = \mathbf{M}_- \boldsymbol{\beta}^k \in \mathbb{R}^N$. Obviously, (9) is fully decentralized as the updates of x_i^{k+1} and α_i^{k+1} only rely on local and neighboring information. Therefore (9) can be used for distributed average consensus. We refer to (9) as the CADMM for distributed average consensus.

If we further initialize $\boldsymbol{\beta}^0$ in the column space of \mathbf{M}_-^T (e.g., $\boldsymbol{\beta}^0 = \mathbf{0}$), then $\boldsymbol{\beta}^k$ lies in the column space of \mathbf{M}_-^T and converges to a unique $\boldsymbol{\beta}^*$. We will use this result immediately but postpone its proof to Lemma 4. Note that this implies $\boldsymbol{\alpha}^k$ in (9) converges uniquely to $\boldsymbol{\alpha}^* = \mathbf{M}_- \boldsymbol{\beta}^*$. We also notice an

interesting relation between α and β even though $M_- M_-^T$ is rank deficient.¹

Lemma 2: Given a connected network, if β lies in the column space of M_-^T , then α and β are one-to-one correspondence, i.e., for $\alpha = M_- \beta$ and $\alpha' = M_- \beta'$ where β and β' are in the column space of M_-^T , $\alpha = \alpha'$ if and only if $\beta = \beta'$.

Proof: That $\beta = \beta'$ implying $\alpha = \alpha'$ is straightforward. Consider $\alpha = M_- \beta$ and write $\beta = M_-^T b$ for some $b \in \mathbb{R}^N$. α', β' and b' are similarly defined. Then we have

$$\begin{aligned} \|\alpha - \alpha'\|_2 &= \|M_- M_-^T (b - b')\|_2 \\ &\geq \tilde{\sigma}_{\min}(M_-) \|M_-^T (b - b')\|_2 \\ &= \tilde{\sigma}_{\min}(M_-) \|\beta - \beta'\|_2, \end{aligned}$$

where $\tilde{\sigma}_{\min}(M_-)$ is the smallest nonzero singular value of M_- , whose existence is guaranteed for a connected graph [29]. We therefore have $\beta = \beta'$ if $\alpha = \alpha'$. ■

It is therefore meaningful to define an *initialization condition* for the CADMM. A similar global convergence property for the CADMM is given in Lemma 3.

Initialization condition for the CADMM: x^0 can be any vector in \mathbb{R}^N and α^0 lies in the column space of $M_- M_-^T$.

Lemma 3 (Global convergence of the CADMM): For any x^0 and α^0 satisfying the initialization condition, the CADMM leads to

$$x^k \rightarrow x^* \text{ and } \alpha^k \rightarrow \alpha^* \text{ as } k \rightarrow \infty,$$

where $x^* = \mathbf{1}x_{\text{avg}}$ and $\alpha^* = r - \mathbf{1}x_{\text{avg}}$ which lies in the column space of $M_- M_-^T$ are both unique.

Proof: Global convergence follows from Lemmas 1 and 2 together with the fact that β^k converges to a unique β^* which lies in the column space of M_-^T .

Now taking $k \rightarrow \infty$ in (9) and using the fact that $x_i^* = x_{\text{avg}}$ for $i = 1, 2, \dots, N$, we have

$$\alpha_i^* = r_i - x_{\text{avg}}.$$

Throughout the rest of this paper, we assume that the CADMM, wherever used, is initialized to satisfy the initialization condition. ■

C. Linear convergence of the CADMM

We investigate two properties of the CADMM; the first property is built on global convergence while the second considers the rate of convergence.

Define $L_+ = \frac{1}{2}M_+ M_+^T$ and $L_- = \frac{1}{2}M_- M_-^T$ which are respectively the signless and signed Laplacian matrices with respect to \mathcal{G}_u . Let $W \in \mathbb{R}^{N \times N}$ be the degree matrix related to the underlying network, i.e., a diagonal matrix with its (i, i) th entry being the degree of node i and other entries being 0. Then $W = \frac{1}{2}(L_+ + L_-)$ and Lemma 2 is an immediate result

¹As defined in Section II-C, $M_- M_-^T = 2L_-$ where L_- is the signed Laplacian matrix of the connected undirected graph and always has 0 as its eigenvalue. See [29].

from the property of L_- [29]. We rewrite (9) in the matrix form as

$$\begin{aligned} x^{k+1} &= (I_N + 2\rho W)^{-1}(\rho L_+ x^k - \alpha^k + r), \\ \alpha^{k+1} &= \alpha^k + \rho L_- x^{k+1}, \end{aligned}$$

or equivalently,

$$s^{k+1} = D s^k, \quad (10)$$

with

$$s^k = \begin{bmatrix} x^k \\ \alpha^k \\ r^k \end{bmatrix},$$

and

$$D = \begin{bmatrix} \rho D_0 L_+ & -D_0 & D_0 \\ \rho^2 L_- D_0 L_+ & I_N - \rho L_- D_0 & \rho L_- D_0 \\ \mathbf{0}_N & \mathbf{0}_N & I_N \end{bmatrix}, \quad (11)$$

where $\mathbf{0}_N$ denotes the $N \times N$ matrix with all entries being 0, $D_0 = (I_N + 2\rho W)^{-1}$, $r^0 = r$, and hence, $r^k = r$. From (10), we have

$$s^k = D^k s^0.$$

It is thus interesting to investigate how D^k behaves as $k \rightarrow \infty$. From (11), a logical approach is to study D^k through the structures of L_- , L_+ and W ; fortunately, the global convergence property of the CADMM provides a simple argument to obtain a rough estimate of D^∞ , which, nevertheless, is good enough for our purpose in establishing the main results. Note that we also have $D^* = D^\infty$ and $s^* = [x^*; \alpha^*; r^*] = [x^\infty; \alpha^\infty; r^\infty] = s^\infty$ as our optima due to global convergence of the CADMM. Our result about D^* is given below.

Theorem 1: Consider D defined in (11). Then

$$D^* = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_N & a_1 \mathbf{1}^T & \frac{1}{N} \mathbf{1} \mathbf{1}^T \\ \mathbf{0}_N & a_2 \mathbf{1}^T & I_N - \frac{1}{N} \mathbf{1} \mathbf{1}^T \\ \mathbf{0}_N & \mathbf{0}_N & I_N \end{bmatrix}$$

for fixed $a_1, a_2 \in \mathbb{R}^N$.

Proof: By Lemma 3, we have for any s^0 that satisfies the initialization condition,

$$s^\infty = \begin{bmatrix} x^\infty \\ \alpha^\infty \\ r^\infty \end{bmatrix} = \begin{bmatrix} x^* \\ \alpha^* \\ r^* \end{bmatrix} = \begin{bmatrix} \mathbf{1}x_{\text{avg}} \\ r - \mathbf{1}x_{\text{avg}} \\ r \end{bmatrix}.$$

Recall that $s^\infty = D^\infty s^0$. If we fix α^0 and r^0 , global convergence implies that $s^\infty = s^*$ regardless of the initial value x^0 . Thus $D_{i1} = \mathbf{0}_N$, $i = 1, 2, 3$. Similarly, fixing x^0 and r^0 , we must have $D_{12}\alpha^0 = D_{22}\alpha^0 = \mathbf{0}$. Since α^0 is initialized in the column space of $M_- M_-^T = 2L_-$ where L_- is the signed Laplacian matrix of a connected undirected graph, D_{12} and D_{22} must be respectively the products of some vectors a_1 and a_2 in \mathbb{R}^N multiplying $\mathbf{1}^T$, such that $D_{12}L_- = D_{22}L_- = \mathbf{0}$. Knowing the form of D_{j1} and D_{j2} , $j = 1, 2$, we see that x^∞ and α^∞ only depend on $r^0 = r$. Together with the facts that $x^\infty = x^*$ has each entry of itself reaching the data average $x_{\text{avg}} = \frac{1}{N}r^T \mathbf{1}$ and that $\alpha^\infty = r - \mathbf{1}x_{\text{avg}}$ for any r , we validate D_{13} and D_{23} as given in the theorem. The remaining blocks, D_{32} and D_{33} , follow directly from the matrix multiplication. ■

Given global convergence, we now turn our attention to the rate of convergence of the CADMM. Recent work of [25], [26] has established the linear convergence of the ADMM. Unfortunately, their results do not apply to the CADMM as their conditions are not satisfied here. In [25], the step size of the dual variable update, i.e., ρ in the λ -update of (8), need be sufficiently small while our CADMM has a fixed step size ρ that can be any positive number (see Remark 3 for further discussion on the choice of ρ). The linear convergence in [26] is established provided that either $g(z)$ is strongly convex or B is full row-rank in (4). In our formulation, however, $g(z) = 0$ is not strongly convex and $B = [-I_{2E}; -I_{2E}]$ is row-rank deficient. Nevertheless, we first give Lemma 4 with regards to the convergence rate of a vector concatenating z and β . A more general result can be found in [28, Theorem 1]. Our proof is similar to that of [28] but simpler.

Lemma 4 ([28, Theorem 1]): Consider the ADMM iteration (8) that solves (6). Define

$$u = \begin{bmatrix} z \\ \beta \end{bmatrix} \text{ and } G = \begin{bmatrix} \rho I_{2E} & 0_{2E} \\ 0_{2E} & \frac{1}{\rho} I_{2E} \end{bmatrix},$$

where β is the dual variable. If we initialize $z^0 = \frac{1}{2}M_+^T x^0$, $\beta^0 = -\gamma^0$ where γ is the other dual variable and β^0 is in the column space of M_-^T , then for $k = 0, 1, \dots$, $z^k = \frac{1}{2}M_+^T x^k$, β^k lies in the column space of M_-^T , and (x^k, z^k, β^k) converges uniquely to (x^*, z^*, β^*) with $x^* = 1x_{\text{avg}}$, $z^* = \frac{1}{2}M_+^T 1x_{\text{avg}}$ and β^* being a vector in the column space of M_-^T . Furthermore, $u^k = [z^k; \beta^k]$ converges Q-linearly to its optimal $u^* = [z^*; \beta^*]$ with respect to the G -norm

$$\|u^{k+1} - u^*\|_G^2 \leq \frac{1}{1+\delta} \|u^k - u^*\|_G^2, \quad (12)$$

where

$$\delta = \min \left\{ \frac{\tilde{\sigma}_{\min}^2(M_-)}{2\sigma_{\max}^2(M_+)}, \frac{4\rho\tilde{\sigma}_{\min}^2(M_-)}{\rho^2\sigma_{\max}^2(M_+)\tilde{\sigma}_{\min}^2(M_-) + 8} \right\},$$

$\sigma_{\max}(M_+)$ denotes the spectral norm or the largest singular value of M_+ , and $\tilde{\sigma}_{\min}(M_-)$ denotes the smallest positive singular value of M_- .

Proof: See Appendix. ■

With this lemma, we can now establish the linear convergence rate of the CADMM.

Theorem 2 (Linear convergence of the CADMM): Consider the matrix form of the CADMM in (10). If $s^0 = [x^0; \alpha^0; r^0]$ satisfy the initialization condition, then

$$\begin{aligned} \|s^{k+1} - s^*\|_2 &= \|(D^{k+1} - D^*)s^0\|_2 \\ &\leq \left(1 + \sqrt{\frac{\rho}{1+\delta}}\sigma_{\max}(M_-)\right) \|u^k - u^*\|_G, \end{aligned}$$

where δ and u^k are defined in Lemma 4. Therefore, s^k is R-linearly convergent to s^* .

Proof: Notice that the initializations in Lemma 4 decentralize the ADMM iteration (8) into the CADMM. Thus x^k is the same in the ADMM iteration (8) and the CADMM iteration (9) while $\alpha^k = M_- \beta^k$. Then (42) implies

$$\|x^{k+1} - x^*\|_2 \leq \|u^k - u^*\|_G.$$

We also have

$$\begin{aligned} \|\alpha^{k+1} - \alpha^*\|_2 &= \|M_- (\beta^{k+1} - \beta^*)\|_2 \\ &\leq \sigma_{\max}(M_-) \|\beta^{k+1} - \beta^*\|_2 \\ &\leq \sqrt{\rho}\sigma_{\max}(M_-) \|u^{k+1} - u^*\|_G \quad (13) \\ &\leq \sqrt{\frac{\rho}{1+\delta}}\sigma_{\max}(M_-) \|u^k - u^*\|_G, \end{aligned}$$

where the last two inequalities are from the definitions of u and G , and (12), respectively. Thus,

$$\begin{aligned} \|s^{k+1} - s^*\|_2 &\leq \|x^{k+1} - x^*\|_2 + \|\alpha^{k+1} - \alpha^*\|_2 \\ &\leq \left(1 + \sqrt{\frac{\rho}{1+\delta}}\sigma_{\max}(M_-)\right) \|u^k - u^*\|_G. \end{aligned}$$

■

III. QUANTIZED CONSENSUS

To model the effect of quantized communications, we assume that each agent can store and compute real values with infinite precision; however, an agent can only transmit quantized data through the channel which are received by its neighbors without any error. The quantization operation is defined as follows. Let $\Delta > 0$ be a given quantization resolution and define the quantization lattice in \mathbb{R} by

$$\Lambda = \{t\Delta : t \in \mathbb{Z}\}.$$

A quantizer is a function $Q : \mathbb{R} \rightarrow \Lambda$ that maps a real value to some point in Λ . Among all quantizers we consider the following two for distributed average consensus:

- 1) Probabilistic quantizer Q_p defined as follows: for $y \in [t\Delta, (t+1)\Delta)$,

$$Q_p(y) = \begin{cases} t\Delta, & \text{with probability } t+1 - \frac{y}{\Delta}, \\ (t+1)\Delta, & \text{with probability } \frac{y}{\Delta} - t. \end{cases} \quad (14)$$

- 2) Rounding quantizer Q_d which projects $y \in \mathbb{R}$ to its nearest point in Λ :

$$Q_d(y) = t\Delta, \text{ if } \left(t - \frac{1}{2}\right)\Delta \leq y < \left(t + \frac{1}{2}\right)\Delta. \quad (15)$$

We point out that probabilistic quantization is equivalent to a dithered quantization method (see [17, Lemma 2]) while rounding quantization is one of the deterministic quantization schemes. Throughout the rest of this paper, we use $Q(y)$ (or $y_{[Q]}$ for ease of presentation) to denote the quantized value of $y \in \mathbb{R}$ regardless of its quantization scheme; we use $Q_p(y)$ (or $y_{[Q_p]}$) and $Q_d(y)$ (or $y_{[Q_d]}$) when it is necessary to specify the quantization scheme. Quantizing a vector means quantizing each of its entries. Define $e_Q = y_{[Q]} - y$ as the quantization error. It is clear that

$$|e_{Q_p}| \leq \Delta \text{ and } |e_{Q_d}| \leq \frac{1}{2}\Delta, \text{ for any } y \in \mathbb{R}. \quad (16)$$

As seen from Section II, the CADMM has the advantage of global and linear convergence for solving the average consensus problem as long as the initialization condition is met. The authors in [15], [16] have also shown the good behavior of the ADMM in distributed settings when noise or random link failures are imposed. The rest of this paper

is devoted to investigating the effects of the two quantization schemes defined in (14) and (15) on the performance of the CADMM. We remark that the results of probabilistic and rounding quantizations hold respectively for other dithered and deterministic cases, which will be elaborated in Sections IV and V.

IV. PROBABILISTIC QUANTIZATION

For ease of presentation, we only study the probabilistic quantization defined in (14). The results can be easily extended to any other dithered quantization as the only information used is the first and second order moments of the probabilistic quantizer output which are stated in the following lemma. See [27] for a proof.

Lemma 5 ([27, Lemma 2]): For every $y \in \mathbb{R}$, it holds that

$$\mathbb{E}[Q_p(y)] = y \text{ and } \mathbb{E}[(y - Q_p(y))^2] \leq \frac{\Delta^2}{4}.$$

The iteration in (9) now takes the form of

$$\begin{aligned} x_i^{k+1} &= \frac{1}{1 + 2\rho|\mathcal{N}_i|} \left(\rho|\mathcal{N}_i|x_{i[Q_p]}^k + \rho \sum_{j \in \mathcal{N}_i} x_{j[Q_p]}^k - \alpha_i^k + r_i \right), \\ \alpha_i^{k+1} &= \alpha_i^k + \rho \left(|\mathcal{N}_i|x_{i[Q_p]}^{k+1} - \sum_{j \in \mathcal{N}_i} x_{j[Q_p]}^{k+1} \right). \end{aligned} \quad (17)$$

Notice that x_i^k is also quantized at its own node for the $(k+1)$ th update; the reason will be given in Remark 5. As illustrated in [15], iteration (17) can be interpreted as a stochastic gradient update. Viewed from this point, the quantization error causes x_i^k to fluctuate around the quantization-free updates (9). Our convergence claims are given in Theorem 3.

Theorem 3: Let \mathbf{x}^0 and $\boldsymbol{\alpha}^0$ satisfy the initialization condition. The probabilistically quantized CADMM (PQ-CADMM) iteration (17) generates $x_i^k, i = 1, 2, \dots, N$, which converges linearly to the data average x_{avg} in the mean sense as $k \rightarrow \infty$. In addition, the variance of x_i^k converges to a finite value which depends on Δ and the network topology.

Proof: Taking expectation of both sides of (17), we have

$$\begin{aligned} \mathbb{E}[x_i^{k+1}] &= \frac{1}{1 + 2\rho|\mathcal{N}_i|} \left(\rho|\mathcal{N}_i|\mathbb{E}[x_{i[Q_p]}^k] + \rho \sum_{j \in \mathcal{N}_i} \mathbb{E}[x_{j[Q_p]}^k] - \mathbb{E}[\alpha_i^k] + r_i \right), \\ \mathbb{E}[\alpha_i^{k+1}] &= \mathbb{E}[\alpha_i^k] + \rho \left(|\mathcal{N}_i|\mathbb{E}[x_{i[Q_p]}^{k+1}] - \sum_{j \in \mathcal{N}_i} \mathbb{E}[x_{j[Q_p]}^{k+1}] \right). \end{aligned} \quad (18)$$

Noting that Lemma 5 implies $\mathbb{E}[x_{i[Q_p]}^k] = \mathbb{E}[x_i^k]$ and $\mathbb{E}[x_{j[Q_p]}^k] = \mathbb{E}[x_j^k]$, we see that (18) takes exactly the same iterations in the mean sense as the CADMM. By initializing $\boldsymbol{\alpha}^0$ in the column space of \mathbf{L}_- , $\mathbb{E}[\boldsymbol{\alpha}^0] = \boldsymbol{\alpha}_0$ satisfies the initialization condition. The linear convergence of $\mathbb{E}[x_i^k]$ to x_{avg} is thus ensured due to Theorem 2.

Since Lemma 5 also indicates the bounded variance of quantization error, the second claim follows directly from [15, Proposition 3]. ■

We notice that the convergence of $\mathbb{E}[x_i^k] \rightarrow x_{\text{avg}}$ does not indicate that \mathbf{x}^k reaches a consensus when $k \rightarrow \infty$. Nevertheless, a simple method fixes this problem. The idea is to calculate the running average $\bar{x}_i^k = \frac{1}{k} \sum_{l=1}^k x_i^l, k \geq 1$ at each node i . One can use similar steps in the proof of [15, Proposition 3] to show that \bar{x}_i^k has diminishing variance. By Chebyshev's inequality, we then get the following corollary.

Corollary 1: Let $\bar{x}_i^k = \frac{1}{k} \sum_{l=1}^k x_i^l$ for $k \geq 1$. For each node i , we have

$$\mathbb{P} \left[\lim_{k \rightarrow \infty} \bar{x}_i^k = x_{\text{avg}} \right] = 1.$$

V. DETERMINISTIC QUANTIZATION

Deterministic quantization is usually much harder to handle as the quantization error is not stochastic. Unlike probabilistic quantization, the accumulated error term is very likely to blow up; there have been a few methods proposed to counter such difficulties (see [11], [19], [20]), yet the resulting algorithms either do not guarantee a consensus or reach a consensus with an error from the desired average that depends on the number of agents, the quantization resolution, and the agents' data. Our approach will establish a finite upper bound on the accumulated error term and then use the property and the initialization condition of local Lagrangian multipliers to deduce the consensus reaching result.

Let the local data x_i^k be also quantized for the $(k+1)$ th update at node i . The updates become

$$\begin{aligned} x_i^{k+1} &= \frac{1}{1 + 2\rho|\mathcal{N}_i|} \left(\rho|\mathcal{N}_i|x_{i[Q_d]}^k + \rho \sum_{j \in \mathcal{N}_i} x_{j[Q_d]}^k - \alpha_i^k + r_i \right), \\ \alpha_i^{k+1} &= \alpha_i^k + \rho \left(|\mathcal{N}_i|x_{i[Q_d]}^{k+1} - \sum_{j \in \mathcal{N}_i} x_{j[Q_d]}^{k+1} \right). \end{aligned} \quad (19)$$

Rewrite $x_{i[Q_d]}^k = x_i^k + e_{iQ_d}^k$ with $e_{iQ_d}^k \in [-\Delta/2, \Delta/2]$ according to (15). Then the α_i -update, $i = 1, 2, \dots, N$, is equivalent to

$$\begin{aligned} \alpha_i^{k+1} &= \alpha_i^k + \rho \left(|\mathcal{N}_i|x_i^{k+1} - \sum_{j \in \mathcal{N}_i} x_j^{k+1} \right) \\ &\quad + \rho \left(|\mathcal{N}_i|e_{iQ_d}^{k+1} - \sum_{j \in \mathcal{N}_i} e_{jQ_d}^{k+1} \right), \end{aligned}$$

or written in the matrix form,

$$\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k + \rho \mathbf{L}_- \mathbf{x}^{k+1} + \rho \mathbf{L}_- \mathbf{e}_{Q_d}^{k+1}, \quad (20)$$

where $\mathbf{e}_{Q_d}^k$ denotes the vector concatenating all $e_{iQ_d}^k$. Recalling the ideal CADMM update (10), we have the matrix form of (19) as

$$\mathbf{s}^{k+1} = \mathbf{D}(\mathbf{s}^k + \mathbf{s}_x^k) + \mathbf{s}_\alpha^k \quad (21)$$

where $\mathbf{s}_x^k = [e_{Q_d}^k; \mathbf{0}; \mathbf{0}]$ and $\mathbf{s}_\alpha^k = [\mathbf{0}; \rho \mathbf{L}_- \mathbf{e}_{Q_d}^{k+1}; \mathbf{0}]$. It is important to note that $\mathbf{e}_{Q_d}^k$ is deterministic and hence the update (21) is deterministic. Our main results are stated in the following theorem.

Theorem 4: Consider the deterministically quantized CADMM (DQ-CADMM) iteration (19). Let \mathbf{x}^0 and $\boldsymbol{\alpha}^0$ satisfy the initialization condition for the CADMM. Then there exists

a finite time iteration $k_0 \geq 1$ such that for $k \geq k_0$ all the quantized variable values

- either converge to the same quantization value:

$$x_{1[Qd]}^k = \dots = x_{N[Qd]}^k \triangleq x_{[Qd]}^*,$$

- or cycle around the average x_{avg} with a finite period $T \geq 2$, i.e., $x_{i[Qd]}^k = x_{i[Qd]}^{k+T}$, $i = 1, 2, \dots, N$, and

$$\frac{1}{T} \sum_{l=1}^T x_{1[Qd]}^{k+l} = \dots = \frac{1}{T} \sum_{l=1}^T x_{N[Qd]}^{k+l} \triangleq \bar{x}_{[Qd]}^*. \quad (22)$$

For both convergent and cyclic cases, we have the following error bound for $x_{Qd}^* \in \{x_{[Qd]}^*, \bar{x}_{[Qd]}^*\}$:

$$|x_{Qd}^* - x_{\text{avg}}| \leq \left(\frac{1}{2} + \rho \frac{2E}{N} \right) \Delta, \quad (23)$$

where the upper bound is tight if the DQ-CADMM converges.

Proof: We prove that the DQ-CADMM either converges or cycles after a finite-time iteration and then use this fact to derive the error bounds.

We see from (20) that α^k must lie in the column space of L_- if α^0 is initialized in the column space of L_- . Following (21), we have

$$\begin{aligned} s^k &= D(s^{k-1} + s_x^{k-1}) + s_\alpha^{k-1} \\ &= D(D(s^{k-2} + s_x^{k-2}) + s_\alpha^{k-2}) + Ds_x^{k-1} + s_\alpha^{k-1} \\ &= \dots \\ &= D^k s^0 + \left(\sum_{i=1}^k D^i s_x^{k-i} + \sum_{j=0}^{k-1} D^j s_\alpha^{k-1-j} \right). \end{aligned} \quad (24)$$

The first term is simply the ideal CADMM update which converges to a finite value. We will show that the accumulated error term $\sum_{i=1}^k D^i s_x^{k-i} + \sum_{j=0}^{k-1} D^j s_\alpha^{k-1-j}$ is bounded and hence that s^k is bounded. Notice that $D^i s_x^{k-i}$ is the i th update of the CADMM with the initial value s_x^{k-i} . Let $u_{k-i}^l = [z_{k-i}^l; \beta_{k-i}^l]$ be the vector that concatenates the primal and dual variables in the ADMM iteration (8), with initial values $z_{k-i}^0 = \frac{1}{2} M_+^T e_{Qd}^{k-i}$ and $\beta_{k-i}^0 = \mathbf{0}$ corresponding to $s_x^k = [e_{Qd}^k; \mathbf{0}; \mathbf{0}]$. With G defined in Lemma 4, we obtain

$$\begin{aligned} \|u_{k-i}^0\|_G^2 &= \rho \left\| \frac{1}{2} M_+^T e_{Qd}^{k-i} \right\|_2^2 \\ &\leq \frac{1}{4} \rho \sigma_{\max}^2(M_+) \|e_{Qd}^{k-i}\|_2^2 \\ &\leq \frac{1}{16} \rho N \Delta^2 \sigma_{\max}^2(M_+), \end{aligned}$$

where the last inequality is from (16). Since Theorem 1 indicates the form of D^* , we get $D^* s_x^{k-i} = \mathbf{0}$, i.e., $x_{k-i}^* = \mathbf{0}$ and $\alpha_{k-i}^* = \mathbf{0}$. Therefore, $u_{k-i}^* = [z_{k-i}^*; \beta_{k-i}^*] = \mathbf{0}$ from Lemma 2 and the fact that $z_{k-i}^* = \frac{1}{2} M_+^T x_{k-i}^*$. Noting also

that the initialization z_{k-i}^0 and β_{k-i}^0 meet the condition of Lemma 4, we thus have

$$\begin{aligned} \|D^i s_x^{k-i}\|_2 &= \|(D^i - D^*) s_x^{k-i}\|_2 \\ &\stackrel{(a)}{\leq} \left(1 + \sqrt{\frac{\rho}{1+\delta}} \sigma_{\max}(M_-) \right) \|u_{k-i}^{i-1} - u_{k-i}^*\|_G \\ &\stackrel{(b)}{\leq} \left(1 + \sqrt{\frac{\rho}{1+\delta}} \sigma_{\max}(M_-) \right) \left(\sqrt{\frac{1}{1+\delta}} \right)^{i-1} \\ &\quad \times \frac{1}{4} \Delta \sigma_{\max}(M_+) \sqrt{\rho N}, \end{aligned} \quad (25)$$

where (a) is from Theorem 2 and (b) is due to Lemma 4 together with the fact that $u_{k-i}^* = \mathbf{0}$. Similarly, we have for $j \geq 1$,

$$\begin{aligned} \|D^j s_\alpha^{k-1-j}\|_2 &\leq \left(1 + \sqrt{\frac{\rho}{1+\delta}} \sigma_{\max}(M_-) \right) \left(\sqrt{\frac{1}{1+\delta}} \right)^{j-1} \\ &\quad \times \frac{1}{4} \Delta \sigma_{\max}(M_-) \sqrt{\rho N}, \end{aligned} \quad (26)$$

and when $j = 0$,

$$\|D^j s_\alpha^{k-1-j}\|_2 = \|s_\alpha^{k-1}\|_2 \leq \frac{1}{4} \Delta \sigma_{\max}(M_-) \rho \sqrt{N}. \quad (27)$$

Therefore,

$$\begin{aligned} &\left\| \sum_{i=1}^k D^i s_x^{k-i} + \sum_{j=0}^{k-1} D^j s_\alpha^{k-1-j} \right\|_2 \\ &\leq \sum_{i=1}^k \|D^i s_x^{k-i}\|_2 + \sum_{j=0}^{k-1} \|D^j s_\alpha^{k-1-j}\|_2 \\ &\leq \|s_\alpha^{k-1}\|_2 + \sum_{i=1}^k (\|D^i s_x^{k-i}\|_2 + \|D^i s_\alpha^{k-1-i}\|_2) \\ &\stackrel{(a)}{\leq} \frac{1}{4} \Delta \sigma_{\max}(M_-) \rho \sqrt{N} + \left(1 + \sqrt{\frac{\rho}{1+\delta}} \sigma_{\max}(M_-) \right) \\ &\quad \times \frac{1}{4} \Delta \sqrt{\rho N} (\sigma_{\max}(M_+) + \sigma_{\max}(M_-)) \sum_{i=1}^k \left(\sqrt{\frac{1}{1+\delta}} \right)^{i-1} \end{aligned} \quad (28)$$

where (a) is from (25)-(27). Then (28) must be finite for $k = 1, 2, \dots$, as $\delta > 0$, and thus s^k is bounded. An important fact from (21) is that the update of s^{k+1} and hence s_x^{k+1} is fully determined by $s^k + s_x^k$ due to the deterministic quantization and the CADMM update. Recalling that $\|s_x^k\|_2 = \|e_{Qd}^k\|_2 \leq \frac{\Delta}{2} \sqrt{N}$ and that $s^k + s_x^k = [x_{[Qd]}^k; \alpha^k; r]$ with each entry of $x_{[Qd]}^k$ being a multiple of Δ , each entry of α being a multiple of $\rho \Delta$, and r being fixed, we conclude that there are only finite possible states of $s^k + s_x^k$. Therefore, s^k is either convergent or cyclic with a finite period $T \geq 2$ after a finite-time iteration.

We next consider error bounds for the consensus value. The consensus error may be studied directly by calculating the accumulated error term in (24). However, the bound in (28) is quite loose in general since it results from the worst case. We alternatively derive the error bounds in the respective case using the fact that the DQ-CADMM either converges or cycles.

Convergent case: The convergence of the DQ-CADMM implies that $\mathbf{s}^{k+1} = \mathbf{s}^k$ for $k \geq k_0$, and hence

$$\mathbf{0} = \boldsymbol{\alpha}^{k+1} - \boldsymbol{\alpha}^k = \rho \mathbf{L}_- \mathbf{x}_{[Qd]}^{k+1}.$$

Since \mathbf{L}_- is the Laplacian matrix of a connected graph \mathcal{G}_u , we must have that $\mathbf{x}_{[Qd]}^{k+1}$ reaches a consensus. Now let $x_{[Qd]}^* \in \Lambda$ denote the convergent quantized value. Then $x_{i[Qd]}^\infty = x_{[Qd]}^*$ for $i = 1, 2, \dots, N$, and $x_i^\infty = x_{[Qd]}^* - e_{iQd}^*$. Summing up both sides of (19) from $i = 1$ to N , we have

$$\begin{aligned} \sum_{i=1}^N (1 + 2\rho|\mathcal{N}_i|) (x_{[Qd]}^* - e_{iQd}^*) \\ = \sum_{i=1}^N \left(\rho|\mathcal{N}_i| x_{[Qd]}^* + \rho \sum_{j \in \mathcal{N}_i} x_{[Qd]}^* + r_i \right), \end{aligned}$$

which is equivalent to

$$x_{[Qd]}^* = \frac{1}{N} \sum_{i=1}^N r_i + \frac{1}{N} \sum_{i=1}^N (1 + 2\rho|\mathcal{N}_i|) e_{iQd}^*.$$

Here we use the fact that $\boldsymbol{\alpha}^k$ lies in the column space of \mathbf{L}_- , i.e., $\boldsymbol{\alpha}^k = \mathbf{L}_- \mathbf{b}^k$ where $\mathbf{b}^k \in \mathbb{R}^N$. Then $\sum_{i=1}^N \alpha_i^k = (\mathbf{L}_- \mathbf{b}^k)^T \mathbf{1} = (\mathbf{b}^k)^T (\mathbf{L}_-^T \mathbf{1}) = 0$. Recalling that $x_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N r_i$ and $|e_{iQd}| \leq \frac{\Delta}{2}$, we finally obtain

$$|x_{[Qd]}^* - x_{\text{avg}}| \leq \left(\frac{1}{2} + \rho \frac{2E}{N} \right) \Delta.$$

The following example shows the tightness of this bound in this convergent case. Consider a simple two-node network with $r_1 = -\frac{3}{2}$ and $r_2 = -\frac{7}{2}$. Set both Δ and ρ to be 1. In this case, we have $E = 1$, $N = 2$ and

$$\mathbf{L}_- = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

We start with $x_{1[Qd]}^0 = x_{2[Qd]}^0 = -1$ and $\alpha_1^0 = -\alpha_2^0 = 1$. One can easily check that our initialization condition is met, and $x_{1[Qd]}^k = x_{2[Qd]}^k = -1$ and $\alpha_1^k = -\alpha_2^k = 1, k = 1, 2, \dots$, in the updates of (19). Hence $x_{[Qd]}^* = -1$ and the consensus error is

$$|x_{[Qd]}^* - x_{\text{avg}}| = \frac{3}{2} = \left(\frac{1}{2} + \rho \frac{2E}{N} \right) \Delta.$$

This coincides with the error bound in (23).

Cyclic case: When the DQ-CADMM cycles with a period T , we must have $\mathbf{s}^{k+T} = \mathbf{s}^k$. Thus, for $k \geq k_0$, we have that

$$\mathbf{0} = \boldsymbol{\alpha}^{k+T} - \boldsymbol{\alpha}^k = \rho \mathbf{L}_- \sum_{l=1}^T \mathbf{x}_{[Qd]}^{k+l},$$

and consequently, $\sum_{l=1}^T \mathbf{x}_{[Qd]}^{k+l}$ reaches a consensus, i.e., (22) is true. Now denote

$$\bar{x}_{[Qd]}^* = \frac{1}{T} \sum_{l=1}^T x_{i[Qd]}^{k+l}, i = 1, \dots, n.$$

We then get

$$\left| \bar{x}_{[Qd]}^* - \frac{1}{T} \sum_{l=1}^T x_i^{k+l} \right| \leq \frac{1}{T} \sum_{l=1}^T |x_{i[Qd]}^{k+l} - x_i^{k+l}| \leq \frac{\Delta}{2}. \quad (29)$$

Summing both sides of (19) over one period and dividing the sum by T , we have

$$\frac{1}{T} \sum_{l=1}^T x_i^{k+l} = \frac{1}{1 + 2\rho|\mathcal{N}_i|} \left(2\rho|\mathcal{N}_i| \bar{x}_{[Qd]}^* - \frac{1}{T} \sum_{l=1}^T \alpha_i^{k+l} + r_i \right).$$

Finally, using (29) and following the same steps as in the convergent case we conclude that

$$|\bar{x}_{[Qd]}^* - x_{\text{avg}}| \leq \left(\frac{1}{2} + \rho \frac{2E}{N} \right) \Delta. \quad \blacksquare$$

Remark 1: The result that deterministic quantization may lead the consensus algorithm to either convergent or cyclic cases is also reported in [19]. Similar to theirs, one can use the history of agents' variables, e.g., running average, to achieve asymptotic convergence at each node. Differently, while they can make local variable values close to the true average in cyclic cases without guaranteeing a consensus, our algorithm can reach a consensus but does not make the error arbitrarily small in general.

Remark 2: We shall mention that $x_{[Qd]}^*$ or $\bar{x}_{[Qd]}^*$ need not be unique. This is because, unlike the ideal CADMM, $\|\mathbf{u}^k - \mathbf{u}^*\|_{\mathcal{G}}$ in the DQ-CADMM need not decrease monotonically due to the quantization that occurs on \mathbf{x}^k at each update. Note also that practical consensus value does not necessarily meet the error bound and we usually have smaller errors than (23) in practice (see Fig. 2). We hence expect better consensus when $(\mathbf{x}^0, \boldsymbol{\alpha}^0)$ are initialized closer to the ideal optima, which leads to a two-stage algorithm for quantized consensus in Section VI.

Remark 3: An interesting observation of our main result is the ADMM parameter ρ . While a small ρ indicates a small consensus error bound, the current paper does not quantify how it affects the convergence rate. Here we do not study the optimal selection of ρ but simply set $\rho = 1$. Therefore we do not regard ρ as a factor affecting our algorithm's performance. We refer readers to [22], [28], [30] for detailed discussions on how ρ affects the ADMM's performance.

Remark 4: Theorem 4 for rounding quantization extends straightforward to other deterministic quantizations as the only information used in our proof is the bounded quantization error. In contrast with [8], [11] where the algorithms may fail for some deterministic quantization schemes, e.g., the rounding quantization, our results work for all deterministic quantization schemes as long as a finite quantization error bound is provided.

Remark 5: In both the PQ-CADMM and DQ-CADMM iterations, x_i^k is quantized for the $(k+1)$ th update at node i even though nodes can compute and store real values with infinite precision. The reason is to guarantee that $\boldsymbol{\alpha}^k$ lies in the column space of \mathbf{L}_- and thus the ideal CADMM update in either the PQ-CADMM or the DQ-CADMM [cf. Equation (21)] possesses the linear convergence property given in Theorem 2. If we do not quantize x_i^k at its own node, Theorem 3 still holds due to $\mathbb{E}[x_{i[Qp]}^k] = \mathbb{E}[x_i^k]$ while Theorem 4 may fail.

Remark 6: In the problem reformulation (4), each node i has its local objective function being $\frac{1}{2}(x_i - r_i)^2$ and $\mathbf{1}x_{\text{avg}}$

minimizes the global objective function $f(x)$ which is the sum of the local objectives. To analyze the DQ-CADMM, we first identify the CADMM update in the matrix form as $s^{k+1} = Ds^k$ where D is fixed throughout the iterations. We then write the DQ-CADMM update as the sum of the ideal CADMM update plus an accumulated error term and finally utilize the linear convergence rate of the CADMM [cf. Equations (25) and (26)]. In general, if the local objective functions do not have linear gradients or the linear convergence rate is not guaranteed (e.g., the LASSO is not differentiable and the corresponding CADMM update in this paper's fashion does not converge linearly), then the current proof no longer holds with deterministic quantization.

VI. ADMM BASED ALGORITHM FOR QUANTIZED CONSENSUS

Let us summarize the two quantized versions of the CADMM: the PQ-CADMM converges linearly to the data average in the mean sense, but it does not guarantee a consensus within finite iterations; the DQ-CADMM, on the other hand, either converges to a consensus or cycles with the same mean of quantized variable values over one period at each node after a finite-time iteration, but results in an error from the average.

As discussed in Remark 2, we can first run the PQ-CADMM $2K$ times to obtain $\bar{x}_i = \frac{1}{K} \sum_{k=K+1}^{2K} x_i^k$, which is a reasonable estimate of x_{avg} at node i according to Theorem 3. Here K can be chosen such that $\mathbb{E}[x_{i[Q_p]}^K]$ is close enough to x_{avg} when we have the knowledge of agents' data and the network topology. Otherwise, we can simply set $K = \lceil 10N (\log_{10}(\frac{1}{\Delta} + 1) + 1) \max\{-\log_{10} \rho, 1\} \rceil$ or as large as permitted. Note also that $\bar{\alpha}_i = \frac{1}{K} \sum_{k=K+1}^{2K} \alpha_i^k$ is also a good estimate of $\alpha_i^* = r_i - x_{\text{avg}}$, and that $\bar{\alpha} = [\bar{\alpha}_1; \bar{\alpha}_2; \dots; \bar{\alpha}_N] = \frac{1}{K} \sum_{k=K+1}^{2K} \alpha^k$ satisfies the initialization condition as α^k lies in the column space of L_- . We can therefore run the DQ-CADMM with this \bar{x}_i and $\bar{\alpha}_i$ as initial values. The probabilistically quantized CADMM followed by deterministically quantized CADMM (PQDQ-CADMM) is presented in Algorithm 1.

VII. SIMULATIONS

This section investigates the performance of the DQ-CADMM and the PQDQ-CADMM via numerical examples. Since existing methods with dithered quantization do not guarantee convergence to a consensus in finite iterations, we only compare our algorithms with those that uses deterministic quantization to reach a consensus, i.e., the gossip based method in [20] and the classical method in [11].

A. Performance of the PQDQ-CADMM, the DQ-CADMM, the gossip based method, and the classical method

To construct a connected graph with N nodes and E edges, we first generate a complete graph consisting of N nodes, and then randomly remove $\frac{N(N-1)}{2} - E$ edges while ensuring that the network stays connected. Set $\Delta = 1$ and assume that agents' data have very high variances in large networks, e.g., let $r_i \sim \mathcal{N}(0, N^4)$. Our settings are

Algorithm 1 PQDQ-CADMM for quantized consensus

Require: Initialize $x^0 = \mathbf{0}$, $\alpha^0 = \mathbf{0}$, and $\rho > 0$. Set $K = \lceil 10N (\log_{10}(\frac{1}{\Delta} + 1) + 1) \max\{-\log_{10} \rho, 1\} \rceil$.

1: **for** $k = 0, 1, \dots, 2K - 1$, every node i **do**
2:

$$x_i^{k+1} \leftarrow \frac{1}{1 + 2\rho|\mathcal{N}_i|} \left(\rho|\mathcal{N}_i|x_{i[Q_p]}^k + \rho \sum_{j \in \mathcal{N}_i} x_{j[Q_p]}^k - \alpha_i^k + r_i \right),$$

$$\alpha_i^{k+1} \leftarrow \alpha_i^k + \rho \left(|\mathcal{N}_i|x_{i[Q_p]}^{k+1} - \sum_{j \in \mathcal{N}_i} x_{j[Q_p]}^{k+1} \right).$$

3: **end for**

4: **set** $x^{2K} = \frac{1}{K} \sum_{l=K+1}^{2K} x^l$, $\alpha^{2K} = \frac{1}{K} \sum_{l=K+1}^{2K} \alpha^l$, and $k = 2K$.

5: **repeat**

6: **For** $i = 1, 2, \dots, N$,

$$x_i^{k+1} \leftarrow \frac{1}{1 + 2\rho|\mathcal{N}_i|} \left(\rho|\mathcal{N}_i|x_{i[Q_d]}^k + \rho \sum_{j \in \mathcal{N}_i} x_{j[Q_d]}^k - \alpha_i^k + r_i \right),$$

$$\alpha_i^{k+1} \leftarrow \alpha_i^k + \rho \left(|\mathcal{N}_i|x_{i[Q_d]}^{k+1} - \sum_{j \in \mathcal{N}_i} x_{j[Q_d]}^{k+1} \right).$$

7: **set** $k = k + 1$.

8: **until** a predefined stopping criterion (e.g., a maximum iteration number) is satisfied.

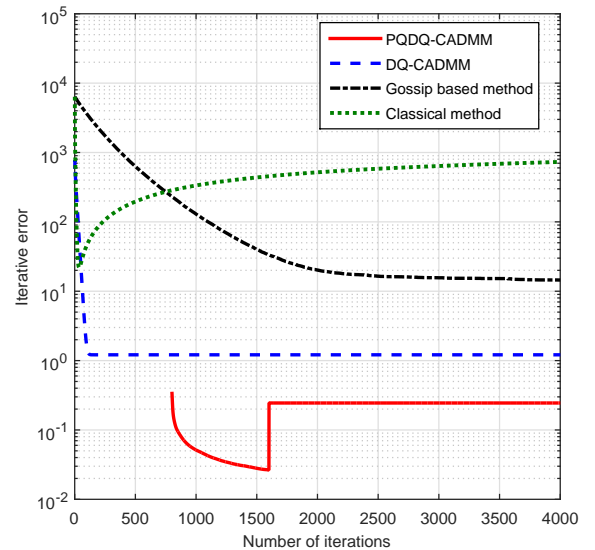


Fig. 1. Iterative error versus iterations where each plotted value is the average of 1000 runs.

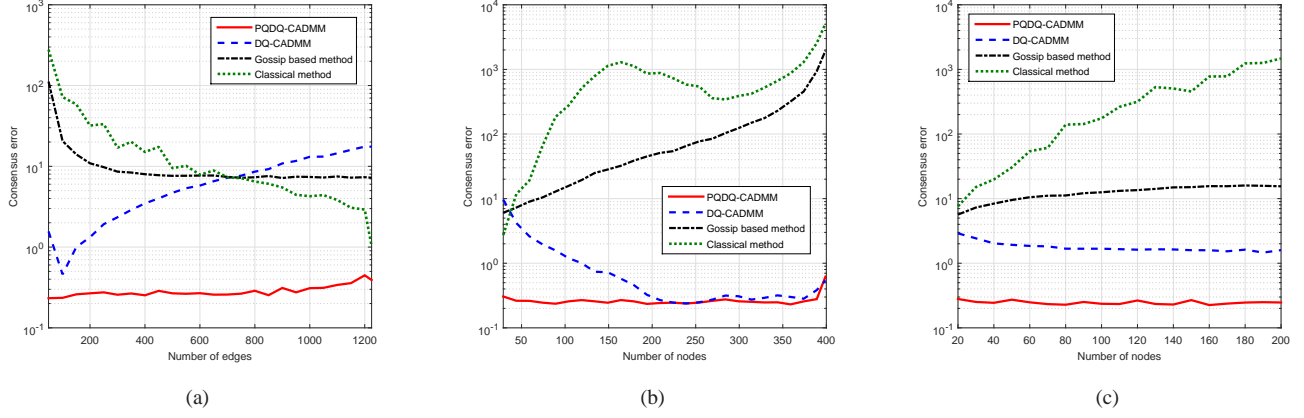


Fig. 2. Consensus error of the four algorithms where $\Delta = 1$ and the plotted values are the average of 100 runs; (a) fixing $N = 50$ and varying $E \in [49, 1225]$, (b) fixing $E = 400$ and varying $N \in [29, 399]$, (c) fixing $\frac{2E}{N} = 10$ and varying $N \in [20, 200]$.

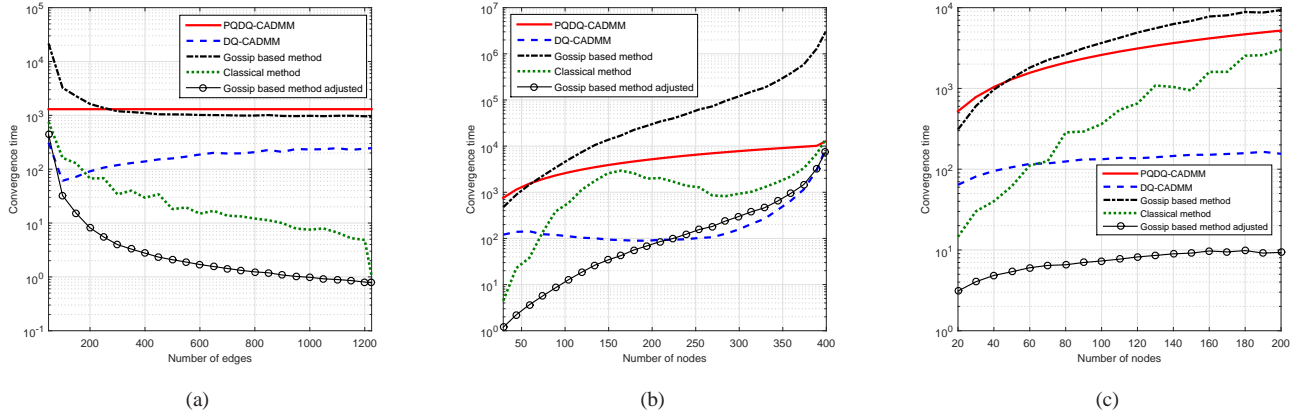


Fig. 3. Convergence time of the four algorithms where $\Delta = 1$ and the plotted values are the average of 100 runs; (a) $N = 50$ and $E \in [49, 1225]$, (b) $E = 400$ and $N \in [29, 399]$, (c) $\frac{2E}{N} = 10$ and $N \in [20, 200]$.

- PQDQ-CADMM: Set $\rho = 1$.
- DQ-CADMM: Set $\rho = 1$, $\mathbf{x}^0 = \mathbf{0}$ and $\alpha^0 = \mathbf{0}$.
- Gossip based method: We randomly pick one edge in \mathcal{A} and perform the updating, i.e., if $(i, j) \in \mathcal{A}$ is chosen, then $x_i^{k+1} = x_j^{k+1} = \frac{1}{2}(x_{i[Q_d]}^k + x_{j[Q_d]}^k)$.
- Classical method: Let \mathbf{W} denote the weight matrix of the graph $\mathcal{G}_d = \{\mathcal{V}, \mathcal{A}\}$. The updating rule is then given by $\mathbf{x}^{k+1} = \mathbf{W}\mathbf{x}_{[Q_{rd}]}^k$ where the subscript $[Q_{rd}]$ denotes the rounding down quantization. We utilize the Metropolis weights defined in [6]:

$$W_{ij} = \begin{cases} (1 + \max\{|\mathcal{N}_i|, |\mathcal{N}_j|\})^{-1}, & (i, j) \in \mathcal{A}, \\ 1 - \sum_{k \in \mathcal{N}_i} W_{ik}, & i = j, \\ 0, & \text{otherwise.} \end{cases}$$

We simulate a connected network with $N = 50$ nodes and $E = 500$ edges. Define the iterative error as $\|\mathbf{x}_{[Q]}^k - \mathbf{1}x_{\text{avg}}\|_2/\sqrt{N}$ which is equal to the consensus error $|x_{[Q]}^* - x_{\text{avg}}|$ when consensus is reached. Plotted in Fig. 1 is the iterative error of the four algorithms at every iteration k with each value being the average of 1000 runs. Note that we start the plot of the PQDQ-CADMM from the $(K+1)$ th iteration as its first K iterations are used only to reach a

neighborhood of x_{avg} ; at the $(2K+1)$ th iteration, $\mathbf{x}_{[Q]}^{2K+1}$ is updated based on the running average of the $(K+1)$ th iteration to the $2K$ th iteration. The figure indicates that all the four algorithms converge to a consensus at one of the quantization levels. The average consensus error of the DQ-CADMM is 1.21, which is much smaller than the upper bound $(\frac{1}{2} + \frac{2E}{N})\Delta = 20.5$. One can also see that the PQDQ-CADMM converges almost immediately after the $2K$ th iteration. In the following we compare the consensus error and the convergence time of the four algorithms via simulations that respectively fix the number of nodes, the number of edges, and the average degree of the graph.

Consensus error: In Fig. 2(a) we fix $N = 50$ and vary E until the graph is complete. The gossip based method and the classical method have decreasing consensus errors as E increases. The consensus error of the DQ-CADMM, however, becomes larger as the average degree and therefore the error bound increase. The PQDQ-CADMM has the smallest consensus error whose average of 100 runs is less than 0.40 for all E . We then fix $E = 400$ and let N vary. Fig. 2(b) shows that the gossip based method and the classical method have increasing consensus errors as N increases. The consensus

error of the DQ-CADMM, on the contrary, decreases when N becomes larger. The PQDQ-CADMM also has the smallest consensus error in this case. In the last setting we fix the average degree $\frac{2E}{N} = 10$ while varying N . The classical method and the gossip based method then both have increasing consensus errors when N and thus the range of agents' data increase. The consensus error of the DQ-CADMM is relatively small compared with the upper bound $(0.5 + \frac{2E}{N})\Delta = 10.5$ and decreases when N becomes larger. The proposed PQDQ-CADMM still has the smallest consensus error whose average of 100 runs is less than 0.2 for all N .

We conclude that the consensus error of the gossip based method and the classical method depends on the average degree of the graph as well as the range of agents' data. Note that their consensus errors can be extremely large for a sparsely connected graph. The DQ-CADMM has an increasing consensus error when the average degree increases while the PQDQ-CADMM performs almost the same for all network structures in terms of the consensus error.

Convergence time: We study the convergence time of the four algorithms via numerical examples in Fig. 3. Since the gossip based method involves only one edge and the other three methods utilize all the edges at each iteration, we plot also the quotient of the convergence time of the gossip based method divided by the number of edges, namely, Gossip based method adjusted, in the figure.

In Fig. 3(a), the gossip based method and the classical method converge slower as the graph becomes sparser. When the average degree is fixed, they have longer convergence time as N increases. Therefore, the convergence time of the gossip based method and the classical method is also affected by the average degree of the graph and the range of agents' data. Different from the gossip based and classical methods, we see in Fig. 3(a) that the convergence time of the DQ-CADMM increase as the graph becomes denser. In Fig. 3(b) and Fig. 3(c), however, the convergence time also increases while the graph becomes sparser, which is possibly because of the increased distance between starting points and optimal values. Exactly characterizing the convergence time of the DQ-CADMM is beyond the scope of the current paper and will be treated as future work. For the PQDQ-CADMM, we observe that the significant portion of its convergence time is spent on achieving an approximate estimate of x_{avg} , i.e., running the PQ-CADMM with $2K$ iterations. With good starting points, the DQ-CADMM converges almost immediately.

B. Performance of the PQDQ-CADMM with different quantization resolutions

We next consider the effect of the quantization resolution on the PQDQ-CADMM. Fig. 4 plots consensus errors of the PQDQ-CADMM with $N = 50$ and $E \in [49, 1225]$ for $\Delta \in \{0.02, 0.1, 0.5, 2.5\}$. The consensus error tends to increase on the average as the quantization resolution becomes larger, which is not surprising since a coarse quantization indicates a higher loss of information at each update. We then calculate the ratio of the consensus error to the quantization resolution: the plotted values, which are the averages of

100 runs, all lie in $(0.227\Delta, 0.337\Delta)$ and the variances are less than 0.051. Moreover, the convergence time of each quantization resolution has a mean of $(2K + 2.1)$ iterations and a variance less than 0.0008, which coincides with our previous analysis that the PQDQ-CADMM converges immediately after the first $2K$ iterations.

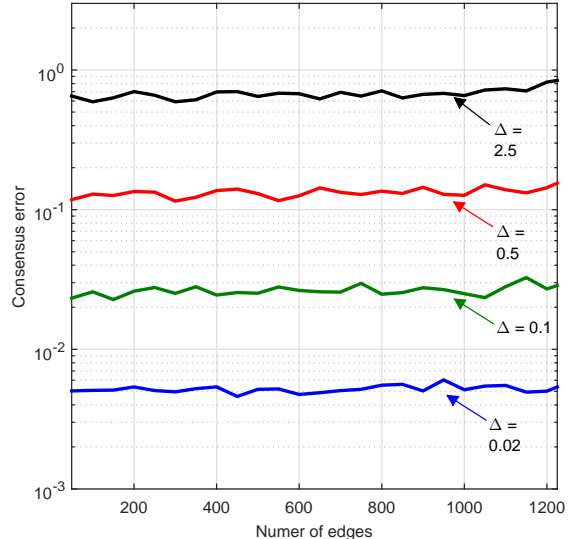


Fig. 4. Consensus error of the PQDQ-CADMM with different quantization resolutions, i.e., $\Delta \in \{0.02, 0.1, 0.5, 2.5\}$, for $N = 50$ and $E \in [49, 1225]$; each plotted value is the average of 100 runs.

C. Cyclic Cases

While we prove that the DQ-CADMM either converges or cycles in Theorem 4, it is noted that the above numerical examples all lead to reach convergence results. Indeed, the proposed deterministic algorithms, the DQ-CADMM and PQDQ-CADMM, converges in most cases as shown by the following simulation. For connected networks with N nodes, we consider star graph which has the smallest average degree, randomly generated graph that has intermediate average degree, and complete graph that has the largest average degree. The result is given in Fig. 5 where the y -axis represents the number of cyclic cases in 10^4 trials. Clearly, the DQ-CADMM and PQDQ-CADMM with fixed parameter $\rho = 1$ converge in most cases, particularly with large networks.

VIII. CONCLUSION

In this paper we have proposed an efficient algorithm, the PQDQ-CADMM, for quantized consensus problems. We first study the effects of both probabilistic and deterministic quantizations on the CADMM. With probabilistic quantization, the PQ-CADMM converges linearly to the data average in the mean sense. In the deterministic case, we can bound the sum of the absolute value of each error term caused by quantization using the global and linear convergence of the CADMM and thus prove that the DQ-CADMM either converges or cycles. We finally combine the two quantized versions of the CADMM to obtain the PQDQ-CADMM algorithm, where the PQ-CADMM is used to get an initial estimate of the data

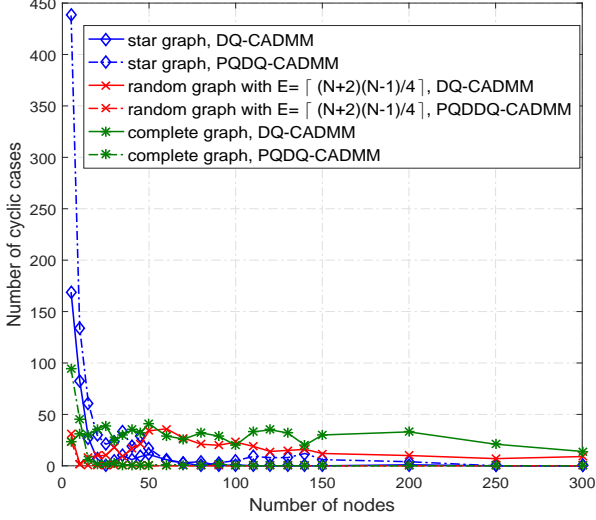


Fig. 5. Number of cyclic cases in 10^4 trials.

average and the DQ-CADMM is used subsequently for consensus reaching purpose. Simulations show that our PQDQ-CADMM provides the best result than all existing methods using deterministic quantization in terms of the consensus error.

Our approach also motivates a number of further research directions:

- 1) Data communications between agents were assumed to be perfect in this paper. In practice, channel impairment may lead to imperfect transmissions. Moreover, the links between agents may fail and the network topology may vary randomly, as studied in [11], [15], [18]. It is thus meaningful to investigate how our algorithm performs in these settings.
- 2) The algorithm parameter ρ is another interesting topic in the DQ-CADMM. Roughly speaking, a smaller ρ may result in a small consensus error but a longer time to reach the convergent or cyclic result. Therefore, its choice should be guided depending on whether a small consensus error or fast consensus speed is desired.
- 3) We only considered the unbounded quantization scheme in this paper. It is also interesting to consider bounded quantization that is used in many applications as it significantly reduces the amount of data that needs to be exchanged.

APPENDIX

Proof of Lemma 4: We first manipulate (8) to derive equivalent updates

$$\nabla f(\mathbf{x}^{k+1}) + \mathbf{A}^T \boldsymbol{\lambda}^k + \rho \mathbf{A}^T \mathbf{B}(\mathbf{z}^k - \mathbf{z}^{k+1}) = \mathbf{0}, \quad (30)$$

$$\mathbf{B}^T \boldsymbol{\lambda}^{k+1} = \mathbf{0}, \quad (31)$$

$$\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k - \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1}) = \mathbf{0}, \quad (32)$$

where (30) and (31) are from multiplying the two sides of the $\boldsymbol{\lambda}$ -update by \mathbf{A}^T and \mathbf{B}^T and adding them to the \mathbf{x} -update and

\mathbf{z} -update, respectively. Recalling $\boldsymbol{\lambda} = [\boldsymbol{\beta}; \boldsymbol{\gamma}]$ with $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^{2E}$ and $\mathbf{B} = [-\mathbf{I}_{2E}; -\mathbf{I}_{2E}]$, we know that $\boldsymbol{\beta}^{k+1} = -\boldsymbol{\gamma}^{k+1}$ from (31). Since we initialize $\boldsymbol{\beta}^0 = -\boldsymbol{\gamma}^0$, we have $\mathbf{B}^T \boldsymbol{\lambda}^k = \mathbf{0}$ for $k = 0, 1, \dots$. Equation (30) then reduces to $\nabla f(\mathbf{x}^{k+1}) + \mathbf{M}_- \boldsymbol{\beta}^{k+1} - \rho \mathbf{M}_+ (\mathbf{z}^k - \mathbf{z}^{k+1}) = \mathbf{0}$, and (32) splits into $\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^k - \rho \mathbf{A}_1 \mathbf{x}^{k+1} + \rho \mathbf{z}^{k+1} = \mathbf{0}$ and $\boldsymbol{\gamma}^{k+1} - \boldsymbol{\gamma}^k - \rho \mathbf{A}_2 \mathbf{x}^{k+1} + \rho \mathbf{z}^{k+1} = \mathbf{0}$. Summing and subtracting these two equations we have $\frac{1}{2} \mathbf{M}_+^T \mathbf{x}^{k+1} - \mathbf{z}^{k+1} = \mathbf{0}$ and $\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^k - \frac{\rho}{2} \mathbf{M}_-^T \mathbf{x}^{k+1} = \mathbf{0}$. With the initialization $\mathbf{z}^0 = \frac{1}{2} \mathbf{M}_+^T \mathbf{x}^0$, $\frac{1}{2} \mathbf{M}_+^T \mathbf{x}^k - \mathbf{z}^k = \mathbf{0}$ holds true for $k = 0, 1, \dots$. Since \mathbf{x}^* is unique and equal to $\mathbf{1}x_{\text{avg}}$ according to Lemma 1, $\mathbf{z}^* = \frac{1}{2} \mathbf{M}_+^T \mathbf{x}^*$ is also unique. To summarize, with the initialization $\boldsymbol{\beta}^0 = -\boldsymbol{\gamma}^0$ and $\mathbf{z}^0 = \frac{1}{2} \mathbf{M}_+^T \mathbf{x}^0$, (30)-(32) reduce to

$$\nabla f(\mathbf{x}^{k+1}) + \mathbf{M}_- \boldsymbol{\beta}^{k+1} - \rho \mathbf{M}_+ (\mathbf{z}^k - \mathbf{z}^{k+1}) = \mathbf{0}, \quad (33)$$

$$\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^k - \frac{\rho}{2} \mathbf{M}_-^T \mathbf{x}^{k+1} = \mathbf{0}, \quad (34)$$

$$\frac{1}{2} \mathbf{M}_+^T \mathbf{x}^{k+1} - \mathbf{z}^{k+1} = \mathbf{0}, \quad (35)$$

which further lead to $\mathbf{x}^k \rightarrow \mathbf{x}^* = \mathbf{1}x_{\text{avg}}$ and $\mathbf{z}^k \rightarrow \frac{1}{2} \mathbf{M}_+^T \mathbf{x}^* = \frac{1}{2} \mathbf{M}_+^T \mathbf{1}x_{\text{avg}}$ uniquely as $k \rightarrow \infty$. Taking $k \rightarrow \infty$ in (33)-(35) and using global convergence, we get

$$\nabla f(\mathbf{x}^*) + \mathbf{M}_- \boldsymbol{\beta}^* = \mathbf{0}, \quad (36)$$

$$\mathbf{M}_-^T \mathbf{x}^* = \mathbf{0}, \quad (37)$$

$$\frac{1}{2} \mathbf{M}_+^T \mathbf{x}^* - \mathbf{z}^* = \mathbf{0}. \quad (38)$$

We can now use (36) to demonstrate the uniqueness of $\boldsymbol{\beta}^*$ if we also initialize $\boldsymbol{\beta}_0$ in the column space of \mathbf{M}_-^T . Note that if $\boldsymbol{\beta}^0$ lies in the column space of \mathbf{M}_-^T then (34) indicates that $\boldsymbol{\beta}^k$ also lies in the column space of \mathbf{M}_-^T , $k = 0, 1, \dots$. The uniqueness of $\boldsymbol{\beta}^*$ then follows from the uniqueness of \mathbf{x}^* and Lemma 2.

Next we show the linear convergence of \mathbf{u}^k . Subtracting (33)-(35) from (36)-(38), respectively, and using $\nabla f(\mathbf{x}) = \mathbf{x} - \mathbf{r}$, we have

$$\mathbf{x}^{k+1} - \mathbf{x}^* = \rho \mathbf{M}_+^T (\mathbf{z}^k - \mathbf{z}^{k+1}) - \mathbf{M}_- (\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^*), \quad (39)$$

$$\frac{\rho}{2} \mathbf{M}_-^T (\mathbf{x}^{k+1} - \mathbf{x}^*) = \boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^k, \quad (40)$$

$$\frac{1}{2} \mathbf{M}_+^T (\mathbf{x}^{k+1} - \mathbf{x}^*) = \mathbf{z}^{k+1} - \mathbf{z}^*. \quad (41)$$

We therefore obtain

$$\begin{aligned} & \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2^2 \\ & \stackrel{(a)}{=} \langle \mathbf{x}^{k+1} - \mathbf{x}^*, \rho \mathbf{M}_+^T (\mathbf{z}^k - \mathbf{z}^{k+1}) - \mathbf{M}_- (\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^*) \rangle \\ & \stackrel{(b)}{=} 2\rho \langle \mathbf{z}^k - \mathbf{z}^{k+1}, \mathbf{z}^{k+1} - \mathbf{z}^* \rangle + \frac{2}{\rho} \langle \boldsymbol{\beta}^k - \boldsymbol{\beta}^{k+1}, \boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^* \rangle \\ & \stackrel{(c)}{=} \|\mathbf{u}^k - \mathbf{u}^*\|_G^2 - \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 - \|\mathbf{u}^k - \mathbf{u}^{k+1}\|_G^2, \end{aligned} \quad (42)$$

where (a) is from (39), (b) is from (40) and (41), and (c) is from the definitions of \mathbf{u} and \mathbf{G} . Due to (42), to prove (12) we only need to show

$$\|\mathbf{u}^k - \mathbf{u}^{k+1}\|_G^2 + \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2^2 \geq \delta \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2,$$

which is equivalent to

$$\begin{aligned} \rho \|z^{k+1} - z^k\|_2^2 + \frac{1}{\rho} \|\beta^{k+1} - \beta^k\|_2^2 + \|x^{k+1} - x^*\|_2^2 \\ \geq \delta \rho \|z^{k+1} - z^*\|_2^2 + \frac{\delta}{\rho} \|\beta^{k+1} - \beta^*\|_2^2. \end{aligned}$$

It then suffices to show

$$\begin{aligned} \rho \|z^{k+1} - z^k\|_2^2 + \|x^{k+1} - x^*\|_2^2 \\ \geq \delta \rho \|z^{k+1} - z^*\|_2^2 + \frac{\delta}{\rho} \|\beta^{k+1} - \beta^*\|_2^2. \quad (43) \end{aligned}$$

The rest of this proof is to establish that $\delta \rho \|z^{k+1} - z^*\|_2^2$ and $\frac{\delta}{\rho} \|\beta^{k+1} - \beta^*\|_2^2$ are upper bounded by two non-overlapping parts of the left side of (43), respectively.

We first have from (41) that

$$\begin{aligned} \|z^{k+1} - z^*\|_2^2 &= \frac{1}{4} \|M_+^T (x^{k+1} - x^*)\|_2^2 \\ &\leq \frac{1}{4} \sigma_{\max}^2(M_+) \|x^{k+1} - x^*\|_2^2. \quad (44) \end{aligned}$$

To upper bound $\|\beta^{k+1} - \beta^*\|_2^2$, we first notice that $\beta^{k+1} - \beta^*$ lies in the column space of M_-^T . Therefore,

$$\|M_- (\beta^{k+1} - \beta^*)\|_2^2 \geq \tilde{\sigma}_{\min}^2(M_-) \|\beta^{k+1} - \beta^*\|_2^2. \quad (45)$$

Now using (45) and (39) we get

$$\begin{aligned} \|\beta^{k+1} - \beta^*\|_2^2 &\leq \frac{1}{\tilde{\sigma}_{\min}^2(M_-)} \|M_- (\beta^{k+1} - \beta^*)\|_2^2 \\ &= \frac{1}{\tilde{\sigma}_{\min}^2(M_-)} \|(x^{k+1} - x^*) - \rho M_+ (z^k - z^{k+1})\|_2^2 \\ &\stackrel{(a)}{\leq} \frac{2}{\tilde{\sigma}_{\min}^2(M_-)} (\|x^{k+1} - x^*\|_2^2 + \rho^2 \|M_+ (z^k - z^{k+1})\|_2^2) \\ &\leq \frac{2}{\tilde{\sigma}_{\min}^2(M_-)} (\|x^{k+1} - x^*\|_2^2 \\ &\quad + \rho^2 \sigma_{\max}^2(M_+) \|z^{k+1} - z^k\|_2^2), \quad (46) \end{aligned}$$

where (a) is from the Cauchy-Schwarz inequality together with the fact $2p_1 p_2 \leq p_1^2 + p_2^2$ for any $p_1, p_2 \in \mathbb{R}$. Combining (44) and (46), we have

$$\begin{aligned} \rho \frac{2\sigma_{\max}^2(M_+)}{\tilde{\sigma}_{\min}^2(M_-)} \|z^{k+1} - z^k\|_2^2 \\ + \left(\frac{1}{4} \rho \sigma_{\max}^2(M_+) + \frac{2}{\rho \tilde{\sigma}_{\min}^2(M_-)} \right) \|x^{k+1} - x^*\|_2^2 \\ \geq \rho \|z^{k+1} - z^*\|_2^2 + \frac{1}{\rho} \|\beta^{k+1} - \beta^*\|_2^2. \end{aligned}$$

The proof is thus complete by picking

$$\delta = \min \left\{ \frac{\tilde{\sigma}_{\min}^2(M_-)}{2\sigma_{\max}^2(M_+)}, \frac{4\rho \tilde{\sigma}_{\min}^2(M_-)}{\rho^2 \sigma_{\max}^2(M_+) \tilde{\sigma}_{\min}^2(M_-) + 8} \right\}.$$

ACKNOWLEDGMENTS

The authors would like to thank Professor Zhi-Quan Luo, from University of Minnesota, Professor Mingyi Hong, from Iowa State University, and Professor Lixin Shen, from Syracuse University, for helpful discussions.

REFERENCES

- [1] S. Zhu and B. Chen, "Distributed average consensus with deterministic quantization: an ADMM approach," in *Proc. IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, Orlando, FL, Dec. 2015.
- [2] W. Ren, R. W. Beard, and E. M. Atkins, "Information consensus in multivehicle cooperative control," *IEEE Control Systems*, vol. 27, no. 2, pp. 71–82, Apr. 2007.
- [3] Y. Cao, W. Yu, W. Ren, and G. Chen, "An overview of recent progress in the study of distributed multi-agent coordination," *IEEE Trans. Ind. Inf.*, vol. 9, no. 1, pp. 427–438, Feb. 2013.
- [4] N. A. Lynch, *Distributed Algorithms*. San Francisco, CA: Morgan Kaufmann, 1996.
- [5] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655–661, May 2005.
- [6] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. Int. Symp. Information Processing in Sensor Networks*, Los Angeles, CA, Apr. 2005.
- [7] C. Xu and F. C. Lau, *Load Balancing in Parallel Computers: Theory and Practice*. Dordrecht, Germany: Kluwer, 1997.
- [8] A. Kashyap, T. Başar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, no. 7, pp. 1192–1203, 2007.
- [9] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 53, no. 1, pp. 65–78, 2004.
- [10] D. Jakovetić, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3788–3801, Jul. 2010.
- [11] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2506–2517, Nov. 2009.
- [12] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, Jul. 2009.
- [13] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.
- [14] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in *ad hoc* WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.
- [15] H. Zhu, G. B. Giannakis, and A. Cano, "Distributed in-network channel decoding," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3970–3983, Oct. 2009.
- [16] T. Erseghe, D. Zennaro, E. Dall'Anese, and L. Vangelista, "Fast consensus by the alternating direction multipliers method," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5523–5537, Nov. 2011.
- [17] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4905–4918, Oct. 2008.
- [18] S. Kar and J. M. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, 2010.
- [19] M. E. Chamie, J. Liu, and T. Başar, "Design and analysis of distributed averaging with quantized communication," in *Proc. 53rd IEEE Conf. Decision and Control*, Los Angeles, CA, Dec. 2014.
- [20] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Gossip consensus algorithms via quantized communication," *Automatica*, vol. 46, no. 1, pp. 70–80, 2010.
- [21] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," DTIC Document, Tech. Rep., 1984.
- [22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [23] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE T. Commun. Techn.*, vol. 12, no. 4, pp. 162–165, Dec. 1964.
- [24] B. He and X. Yuan, "On the $O(1/n)$ convergence rate of the douglas-rachford alternating direction method," *SIAM J. Numer. Anal.*, vol. 50, no. 2, pp. 700–709, 2012.
- [25] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *arXiv preprint arXiv:1208.3922*, 2012.
- [26] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," *J. Sci. Comput.*, vol. 66, no. 3, pp. 889–916, 2016.

- [27] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.
- [28] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the admm in decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750–1761, Apr. 2014.
- [29] F. R. Chung, *Spectral Graph Theory*. American Mathematical Soc., 1997, vol. 92.
- [30] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 644–658, Mar. 2015.